AL-TP-1992-0044

AD-A258 148

ARMSTRONG LABORATORY

NEURAL NETWORK APPLICATIONS:
A LITERATURE REVIEW

DTIC
ELECTE
DEC 1 0 1992
S C D

Vince L. Wiggins

RRC, Incorporated
3833 Texas Avenue, Suite 256
Bryan, TX 77802

Sheree K. Engquist, Major, USAF
Larry T. Looper

HUMAN RESOURCES DIRECTORATE
MANPOWER AND PERSONNEL RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

November 1992

Interim Technical Paper for Period September 1989 - September 1992

393367

92-31185

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

This technical paper is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.
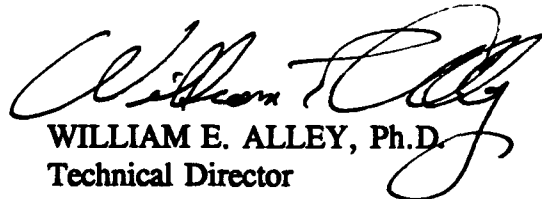
When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.
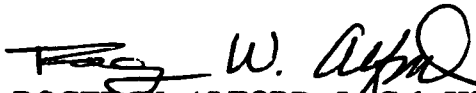
This paper has been reviewed and is approved for publication.

SHEREE K. ENGQUIST, Maj, USAF
Project Scientist

WILLIAM E. ALLEY, Ph.D.
Technical Director
Manpower & Personnel Res Div

ROGER W. ALFORD, Lt Col, USAF
Chief, Manpower and Personnel Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 1992	3. REPORT TYPE AND DATES COVERED Interim - September 1989 – September 1992

4. TITLE AND SUBTITLE

Neural Network Applications: A Literature Review

5. FUNDING NUMBERS

C - F41689-88-D-0251
PE - 62205F
PR - 7719
TA - 20
WU - 20

6. AUTHOR(S)

Vince L. Wiggins
Sheree K. Engquist
Larry T. Looper

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

RRC, Incorporated
3833 Texas Avenue, Suite 256
Bryan, TX 77802

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Manpower and Personnel Research Division
Brooks Air Force Base, TX 78235-5000

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AL-TP-1992-0044

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Captain Sheree Engquist, (210) 536-2257

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Neural networks have been applied to a very broad range of tasks in many different disciplines. The focus of the discussion is on those areas closely related to potential applications in the Air Force personnel system. Similarities between the neural network applications and existing or potential personnel problems were described. Results of other researchers were examined in areas with problems similar to the personnel system and their specific architectures discussed.

14. SUBJECT TERMS

Artificial networks
Neural networks
Personnel research

15. NUMBER OF PAGES

44

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

CONTENTS

CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

PREFACE

v

NEURAL NETWORK APPLICATIONS: A LITERATURE REVIEW

INTRODUCTION

Neural networks have been applied to a very broad range of tasks in many different disciplines. The discussion in this section will focus on those areas most closely related to potential applications in the Air Force personnel system. Where significant, similarities between the neural network applications and existing or potential personnel problems will be briefly pointed out. The application categories below are somewhat arbitrary. Generally, the same characteristics which serve neural networks in classification problems are also useful in prediction or control problems. Despite the wide range of applications, the neural network capabilities being exploited are generally those discussed earlier: ability to perform universal function approximation, ability to form complex decision boundaries from classification examples, and ability to generalize outside of the training data set. These same capabilities are important in virtually all personnel models which are derived from observed behaviors or known system constraints.

The objective of the reviews in this section are twofold. First, the ability of neural networks to perform well on problems similar to those in the personnel system is assessed through the results of other researchers. Second, neural network architectures suitable to personnel problems are identified and their application to specific problems is presented. Several hundred reports have been published in recent years on attempts to apply neural networks to specific problems. Anyone seeking further documentation can consult the proceedings from any of three major neural network conferences held annually in the United States: two Joint International Neural Network Conferences sponsored by the Institute of Electrical and Electronics Engineers (IEEE) and the International Neural Network Society (INNS), and the IEEE Conference on Neural Information Processing Systems - Natural and Synthetic. Collectively, these proceedings document well over 100 new neural network applications each year. In addition, several professional journals are now devoted exclusively to neural networks and publish a considerable number of applications oriented articles. The applications discussed below were chosen on the basis of several criteria: relation to personnel problems, exposition of similarities and differences with traditional methods, evaluation of neural network results against traditional methods, and comparison of results using different neural network architectures.

It should be noted that an enormous corpus of work has also been produced on modelling biological systems with neural networks. Much research has been done on early vision, hearing, and other sensory processing. Another dynamic area of research involves optimization with neural networks. Combinatorial optimization areas such as routing, scheduling, and resource utilization have also been studied extensively. As hardware implementations become available, neural network methods may become the fastest way to solve many of these problems. While such optimization is of interest in the personnel area, optimization speed has rarely been a limiting factor for personnel models, and the hardware for direct implementation of networks

is not yet available. When these hardware solutions become available, they will likely contain "canned" solutions to most typical optimization problems. This review will completely ignore the burgeoning literature in these biological and optimization areas.

CLASSIFICATION APPLICATIONS AND COMPARISONS OF METHODOLOGIES

Analysis of classification problems is one of the most mature areas of neural network research. Classification encompasses a wide range of important tasks in many disciplines. It involves the mapping of a vector of known values into a set of classes or categories. These classes may be distinct and non-overlapping (mammals/reptiles/amphibians) or indistinct and overlapping (plant eaters/meat eaters). They may be specified (animal phyla), behavioral (reenlist/separate), diagnostic (malfunction/normal), predictive (stock prices rising/falling), or any other basis for separating exemplars. The separations may be deterministic, where exemplars which having identical known vectors are always classified into the same class. Or they may be stochastic, where exemplars with identical known input vectors may fall into different classes. In general, the second case is more interesting and involves the extraction of a classification model on the basis of noisy and conflicting examples of the classification. It is also the most common form of classification in the personnel system and many examples can be cited: reenlistment/separation/extension behavior, retirement decisions, job selection for new enlistees, accession decisions (prior and non-prior service), promotion decisions, retraining decisions, pilot weapon systems tracking, choice of flying maneuver, etc.

As of this writing, no research results have been published on the application of neural networks to personnel classification tasks. However, many related areas have been studied and reported. Some of these results demonstrate the capability of networks to perform in the personnel area. They often show comparisons between the performance of neural networks and traditional classification techniques in a particular problem domain. In addition, these studies indicate potential problem areas and aspects of neural networks which have yet to be adequately addressed.

Tests on Contrived Problems

Some of the most important neural network classification tests have been performed on contrived or artificial data sets. In these cases, the researcher(s) builds a data set with known characteristics (decision boundaries, noise levels, etc.) and performs classification tests where the underlying model which generated the exemplars is known. Tests of this sort are important because the behavior and performance of a "best possible" model of the data set is known. While they cannot capture the depth of many "real world" data sets, these tests offer an arena where different methodologies can be compared against a known underlying model. Since current theoretical results on neural networks do not extend to global solutions or generalization performance, these empirical tests provide the only information on these important capabilities.

2

Wave Form Classification

One of the most comprehensive comparative tests was performed by de Bollivier, Gallari, & Thiria (1990) on waveforms. These researchers combined waveforms and added noise to the inputs to produce a difficult, nonlinear classification problem with three classes and twenty-one noisy inputs. While the details of producing the data set are somewhat involved, this data set is highly representative of many personnel classification problems. Details of how the data set was produced can be found in Brieman, Freidman, Rolshen, & Stone. In addition to this base data set, the researchers produced a second, forty-dimension data set by adding nineteen input variables composed purely of random noise. These variables represent superfluous regressors or inputs which are uncorrelated with the desired output (correct classification). This situation is fairly common in any empirical work and particularly in personnel research. Factors which are assumed to be important in classifying a group are often found to have little or no empirical influence on the classification.

De Bollivier et al. tried several traditional and neural network classification techniques on these data sets: discriminant analysis, classification and regression trees (CART), nearest neighbor, K-means, LVQ, and back propagation. In addition, for reference purposes, they computed the results of a Bayes classifier on the problem. Normally, the performance of this classifier is unknown; however, because the researchers knew the exact form of the model and type and quantity of noise, the Bayes result could be analytically derived. Discriminant analysis is well known and probably the most common method for forming linear classification boundaries (see Devijver & Kittler; 1982, Duda & Hart, 1973; or Maddala, 1983 for different treatments). Nearest neighbor and K-means classification are very common classification techniques and were briefly considered in Section II under Learning Vector Quantization. In this case, ten reference means were allocated for each class (a total of thirty means). The two neural network architectures, LVQ and back propagation, are also discussed above under their own headings in Section II. Like K-means, LVQ was allocated a total of thirty reference vectors on the twenty-one input data set. Both LVQ and K-means were allotted thirty-six vectors (or means) for the higher dimensional, forty input problem. The back propagation architecture employed contained two hidden layers. On the twenty-one input problem, the network had twenty-one input neurons, fifteen and nine neurons in two hidden layers, and three output neurons. On the forty input problem, thirty and twenty neurons were allocated to the hidden layers, while the input obviously went to forty and the output remained the same.

CART is a relatively new procedure which is essentially an iterative extension of the linear probability model (Maddala, 1983). First, the data set is divided into two groups using a linear model which best separates the classes. Hyperplanes are formed in the input space to separate the classes. This process is repeated on data in each of the two resulting groups. This procedure is then repeated on the four resulting groups and the process continues until each observation has been split into its own separate sub-sample. Normally, this regression tree is then pruned to its optimal size by examining its performance on a hold-out sample (although, other rules may be used to prune the tree). De Bolivier et al. do not reveal their method for pruning the regression tree.

3

The researchers generated 300 exemplars for training and 5000 for testing the generalization power of each technique. Each technique was trained or estimated to the 300 training exemplars and its performance was evaluated on the 5000 exemplars the method had not seen. The percent of the 5000 testing exemplars that were correctly classified (hit-rate) by each technique is reported in Table 1. The Bayes limit is provided as a reference. Given the level of noise in the data set and the overlap between the classes, the Bayes result is the best separation that is theoretically obtainable on the problem. De Bollivier et al. took some of their results from Brieman et al.; and, since the latter did not consider the "hard" forty input classification problem, these results are not available for all of the techniques. The final two rows of the table show the results of a hybrid architecture combining back propagation and LVQ.

As can be seen in the table, all of the methods except of linear discriminant analysis and, to lesser extent, nearest neighbor perform well on the twenty-one input example where all inputs are meaningful. On this problem, the only "traditional" method whose performance is comparable to the back propagation and LVQ neural networks is K-means analysis. While de Bollivier et al. do not provide a statistical analysis, a simple test on the reported results indicates that the hit-rate for their hybrid algorithm is significantly better than the K-means hit-rate (at better than a 99.9% significance level). In addition, the final hybrid network's performance approaches the theoretical limit for the problem. Using this same test, the performance of K-means, back propagation, and LVQ cannot be distinguished at the 95% level of confidence.

The performance of the techniques on the forty input problem demonstrates a shortcoming of LVQ discussed earlier. As can be seen in Table 1, both LVQ and K-means perform dramatically worse on the forty input problem. This is particularly interesting because the forty input problem contains all of the input information contained in the twenty-one input problem. The only difference is the addition of nineteen inputs who take on random values in each exemplar. LVQ and K-means associate an exemplar with a reference vector by computing an unweighted Euclidean distance between the two vectors. In so doing, they implicitly assume that each input is equally important in performing a classification. This makes these techniques susceptible to superfluous or less important inputs since the techniques cannot "ignore" or discount these inputs. If sufficient reference vectors are allocated and unlimited training data is available, LVQ and K-means can form collections of reference vectors along the superfluous input dimensions. In this manner they can somewhat overcome the problem; but, they are still making inefficient use of the information in the training data set. Back propagation, on the other hand, has no difficulty ignoring the random inputs. Back propagation's performance on the forty input problem is statistically indistinguishable from its performance on the twenty-one input problem.

4

**Table 1. Hit-rate Performance of Various Classification Techniques
on Difficult Nonlinear Waveform Inputs
(Results from de Bollivier et al.)**

Classification Technique	Out of Sample Hit-rate	
	21 Inputs	40 Inputs
Bayes Limit	86.0	86.0
Discriminant Analysis	74.0	-
CART	80.0	-
K-means	82.0	53.2
Nearest Neighbor	78.0	-
LVQ	82.7	53.2
Back Propagation	81.6	80.5
Back Propagation feeding LVQ	83.4	81.2
Hybrid Back Propagation/LVQ	85.0	-

It can also be seen that the "back propagation feeding LVQ" algorithm (a simplified version of the researchers hybrid algorithm) performs well in both cases. By feeding the back propagation network's hidden layer outputs into an LVQ network, the researchers essentially used the back propagation network to filter or weight the inputs before LVQ classification was performed. The other benefit claimed by de Bolliver et al. for the hybrid algorithm over standard back propagation is speed of convergence. On these problems they found the hybrid to converge over ten times faster than standard back propagation.

Multivariate Normals

The performance of three different neural network architectures on another artificial problem was considered by Kohonen, Barna, & Chrisley (1988). In this case the researchers chose a two class problem with two to eight dimensions for the inputs. Each class was generated as a multivariate normal distribution having different means and variances. One problem was designed to be easily separable; and, the means of the two distributions were chosen such that minimal overlap occurred in the distributions. A more difficult problem was also generated in which both distributions had identical means. In this case, the distributions are heavily overlapping; in fact, the distribution with the smaller variance is completely enclosed

5

by the distribution with a larger variance. The optimal decision boundary for this problem must form a hypersphere around the smaller variance distribution at the points where the two distributions intersect. In all, the researchers considered sixteen problems: two through eight dimensional minimal overlapping gaussians for the easy problem, and two through eight dimensional completely overlapping gaussians for the hard problem. In all cases, a total of 1550 samples were drawn from the two distributions as training exemplars and an independent set of 1550 samples were drawn for out-of-sample testing. As with the waveform problem, the known form of the model and error allowed Kohonen et al. to compute the Bayes limit (or optimal classification) for the models.

The three network architectures tested by Kohonen et al. were back propagation, LVQ, and the Boltzman machine (see Ackley, Hinton, & Sejnowski, 1985)[1]. Two Boltzman machine models were considered. In one, the real vector values were used as inputs; in the other, each input dimension was split into twenty segments and a twenty input binary code was used for each input dimension. A detailed discussion of data encodings sometimes used in neural networks is presented in Hancock (1988). The back propagation networks tested all had one hidden layer with eight neurons. Two outputs were used (one for each class), and the number of inputs matches the dimension of the distribution (two to eight). Each LVQ network also had two to eight inputs and two outputs. The LVQ hidden layer contained five reference vector neurons for each dimension in the input (ten to forty neurons).

A summary of Kohonen's et al. results is shown in Table 2. Both LVQ and back propagation can be seen to perform near optimal classification when the classes are represented by two dimensional Gaussians. This result holds for the "easy" case of slightly overlapping distributions and the "hard" case of completely overlapping distributions. When the classes are represented by eight dimensional, slightly overlapping distributions, LVQ performs marginally better than back propagation. The researchers found LVQ's relative performance to be even better when the completely overlapping distributions were tested[2]. In general, the results for three to seven dimensional distributions mirror those shown in the table. As dimension increases, both techniques (but particularly back propagation) decrease their performance compared to the theoretical limit. Kohonen et al. also note that the back propagation results

---

[1] The Boltzman machine combines a hidden or non-hidden layer architecture with a simulated annealing process to find global maxima in probability distributions. Unlike the feed forward networks discussed earlier, it does not distinguish between inputs and outputs (an autoassociative model). It is one of the most computationally intensive neural networks and may require a factor of 100 or 1000 times more processing than a back propagation network. See Ackley, Hinton, and Sejnowski (1985) for implementation details and some applications.

[2] The results for the continuous value Boltzman machine (not displayed) were very poor on the "easy" problem and would not converge on the "hard" problem. The binary value Boltzman results approached the theoretical classification limit in all cases. However, with binary encodings, the problems are really not comparable.

"seemed more unstable" than those for LVQ. Given that LVQ could be trained over one hundred times faster than back propagation, the researchers found LVQ a preferred architecture. RRC Inc's experience with the two architectures indicates that keeping the number of neurons in the hidden layer of the back propagation network constant may have biased the results in favor of LVQ.

Hush & Salas (1990) looked at the same problem using only the eight dimensional case where both gaussians are centered at zero (the "hard" task above). The researchers used two training samples: one with 400 exemplars and a second with 3200 exemplars. They tested four network architectures: back propagation, LVQ, high order neural networks (HONNs), and localized receptive fields. HONNs are similar to standard back propagation networks except they allow high order combinations of neuron outputs in their interconnections between the layers. Hush & Salas used only second-order combinations applied only at the input layer. Because of their structures, both HONNs and localized receptive fields were able to form trivial solutions to the problem. While they achieved optimal performance on this problem, the performance is merely an artifact of the chosen problem. A different problem would have to be considered to make these results interesting.

**Table 2. Hit-rate Performance of Back Propagation and LVQ
on Classes Representing Overlapping Multivariate Gaussians
(Results from Kohonen, Barna, & Chrisley)**

Classification Technique	Slightly Overlapping Distributions		Completely Overlapping Distributions	
	two inputs	eight inputs	two inputs	eight inputs
Bayes Limit	83.7	93.8	73.6	91.0
Back Propagation	83.6	88.7	73.5	81.1
LVQ	83.0	90.0	73.4	86.6

Hush & Salas varied the number of neurons in the hidden layers of LVQ and back propagation between one and one hundred. They found that back propagation's out-of-sample performance degraded when trained with more than thirty-five number of hidden units on the small sample. This is consistent with the over-fitting discussed in Section II. LVQ exhibited this same problem, but the degradation was less noticeable and started at about 60 neurons. Neither technique experienced this over-fitting problem when the larger training sample with

3200 exemplars was used. The approximate out-of-sample hit-rates for the best LVQ and back propagation networks are shown in Table 3.

**Table 3. Performance ⌐f Back Propagation and LVQ
on Classes Representing Eight Dimensional
Completely Overlapping Gaussians
(Results from Hush & Salas)**

Classification Technique	Out of Sample Hit-rate	
	400 Training Exemplars	3200 Training Exemplars
Bayes Limit	91.0	91.0
Back Propagation (30 hidden neurons)	85.0	89.0
Back Propagation (8 hidden neurons)	81.0	82.0
LVQ (40 reference vector neurons)	83.0	86.0

Note: All hit-rates were taken from graphs in Hush & Salas and are approximate.

It is interesting to compare these results to those of Kohonen et al. As can be seen, the back propagation network with eight hidden neurons (the same size as used by Kohonen) performs substantially worse than the back propagation network with thirty hidden neurons. The performance of the smaller back propagation network and the LVQ network are very similar to those of Kohonen et al (Kohonens training sample fell between the two Hush & Salas samples in size). Hush & Salas found LVQ to perform best on this problem when forty to eighty reference vector neurons were available. Kohonen's use of forty neurons falls in this window. However, Hush & Salas found back propagation to perform best with thirty to sixty neurons and required twenty to even approach this level of performance. Thirty to sixty represents best performance on the large training sample, twenty-five to thiry-five neurons performedbest on the smaller (400 exemplar) sample The eight neurons employed by Kohonen et al. were simply too few to capture the complexity of the eight dimensional problem. Hush & Salas' findings indicate that a well chosen back propagation network outperforms a well chosen LVQ network on this problem. In addition, their findings reenforce the need for additional techniques to improve generalization capabilities of neural networks (as discussed in Section II). This type

of generalization behavior as network size is varied is not restricted to artificial problems; most "real-world" data sets exhibit similar problems.

Signal Detection

Gallinari, Thiria, & Soulie (1988) tested back propagation and linear discriminant analysis on a problem that essentially involves signal detection. An eight bit input stream was presented and the network or discriminant function was to detect any cases where three or more bits were "on" in the stream. The researchers first demonstrated a parallel between back propagation and discriminant analysis. They showed that an adaptation of the back propagation network with linear activation functions and one hidden layer of neurons performs discriminant analysis[3]. They found that the nonlinear decision surface provided by the back propagation network allowed for substantial classification improvements. On the 150 exemplars in the training set, discriminant analysis could correctly classify 88%. A three layer back propagation network, with three neurons in the hidden layer, could correctly classify 99% of the training data set. When tested on 106 new exemplars, discriminant analysis correctly classified 76% while back propagation correctly classified 87%.

Other Contrived Tests

Huang and Lippman (1987) ran tests of back propagation, K-nearest neighbor, and Gaussian (Duda & Hart, 1973) classifiers on several one dimensional input problems. K-nearest neighbor algorithms are derived from the nearest neighbor algorithm outlined in Section II. Instead of assuming a new exemplar behaves like its nearest neighbor in the training set, K-nearest neighbor algorithms take a "vote" of the closest K training exemplars and classify the new exemplar with the majority. The best choice for K is problem specific and depends on the size of the training set, underlying model, and noise level. Huang and Lippman found that the K-nearest neighbor and back propagation techniques were more robust to outliers and skewed distributions.

While no tests were made against other methods, Lang & Witbrock (1988) performed an impressive demonstration of back propagation's ability to solve a highly nonlinear classification problem. As seen in Figure 1, two intertwined spirals are assumed to emanate from the origin of a two dimensional space. Each spiral represents a class which can be characterized by its position in the two dimensional space. Because the spirals are intertwined, the classification problem is particularly difficult. Over most of the problem range, each class is bounded on all four sides by the other class and then picks up again on the other side of the boundary. The decision boundaries between the two regions are extremely complex. Using a network with two

---

[3]Back propagation networks have been shown to be able to reproduce several "standard" statistical techniques. Oja (1982) showed that a network with linear activation functions and trained to reproduce its own inputs could perform extract the first principle component from a data set. This result was extended to higher principle components by Hrycej (1990).
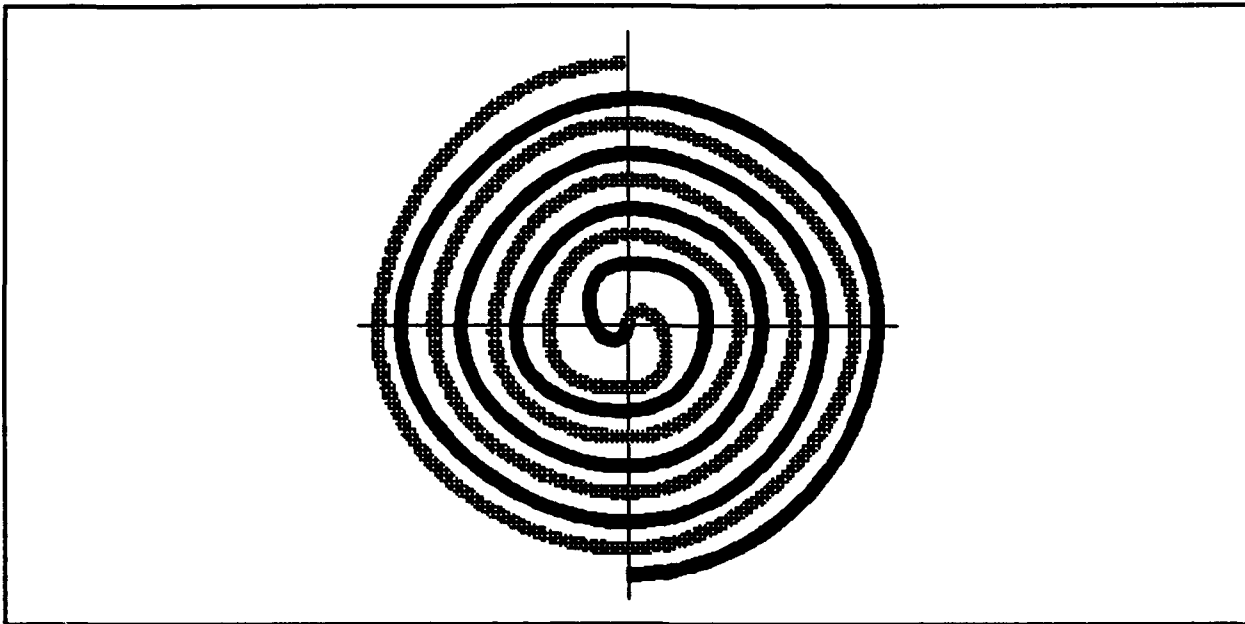
Figure 1. Intertwined spiral classification test. Dark and grey lines form separate classes.

inputs (the x and y coordinates) and three hidden layers (each containing five neurons), Lang & Wittbrock were able to form this decision boundary using only 191 training points (exemplars)[4]. While one is very unlikely to encounter such a complex decision boundary in any actual data set, this demonstration places a high upper bound on the ability of back propagation to discern complex boundaries.

Financial Applications

Several applications and tests of neural networks on problems in the financial sector have been reported. Typical examples include bond rating, stock market prediction, and assessment of mortgage applicants. While not directly comparable to personnel classification problems, these examples share several features with personnel classification. Bond rating, for example, requires that a company be assessed based on its past performance characteristics. This process is not dissimilar from the assessment the Air Force must make when choosing accession applicants or UPT candidates. Selection and rejection of mortgage applicants is even closer to these Air Force personnel decisions. In this case, individuals are assessed on their

---

[4]Unlike most back propagation networks, Lang and Witbrock completely connected each layer to the neurons in all preceding layers. For example, each neuron in the third hidden layer was completely connected to all neurons in the first hidden layer and the two input neurons as well as the usual connections to the second hidden layer. They found the increased flexibility of this architecture necessary to solve this particular problem.

characteristics, current financial status, and financial history. While this is a very active area of research, much of the work in this area has not been published. Companies are hesitant to release information which may give them an edge in the market place.

Bond Rating

Surkan & Singleton (1990) have compared the ability of back propagation networks and linear discriminant analysis, a widely applied and accepted method of classification in financial research, to reproduce the bond ratings produced by such companies as Standard & Poors or Moodys. Their training and testing data sets were drawn from a fairly homogenous set of companies -- the eighteen companies divested by ATT in 1982. The researchers chose to aggregate the many possible bond classifications into two groups: Aaa bonds and A1 through A3 bonds. Aaa bonds are the highest quality bonds with A1, A2, and A3 forming the next lower quality tier or "investment grade" bonds. Seven common financial ratios and rates of returns for the companies were used to classify the companies into the two bond rating categories. These financial indicators included such values as return on equity, construction expenditures over cash flow, and the log of total assets. All data was taken from bonds issued by the companies from 1982 through 1988 and consisted of fifty-six bond issues. These issues were divided into a very small training set of sixteen issues (ten Aaa and six A1-3) and forty testing issues (twenty in each rating class).

Three different network architectures were trained to the sixteen training exemplars. Two of the networks employed two hidden layers and one used a single hidden layer (the number of neurons can be seen in Table 4). The target for training the networks (or discriminant analysis) were the actual ratings of Standard & Poors or Moodys for the issues. Because of the small training sample, Surkan and Singleton employed an unusual sequencing method for presenting exemplars during training. They randomly sampled exemplars from each of the two classes without replacement and alternated selections from each class. In this manner an equal number of presentations were made from each class, despite the fact the Aaa training class contained sixty percent more training exemplars[5]. As usual, the networks were trained to convergence on the sixteen observation training set, and the resulting networks were tested on the forty observation hold-out sample.

A different testing method was employed by the researchers to assess the performance of the discriminant analysis. In this case, they used a hold-one-out process of testing the ability of dicriminant analysis to predict observations on which it had not been estimated. Application of hold-one-out sampling was straight forward. The discrimimant analysis was estimated on fifty-five of the fifty-six bond issues. The resulting discriminant model was used to predict the class of the single issue not in the estimation sample. This process was repeated until each of

---

[5]Other researchers such as Lippman (1987) and the aforementioned Hoskins (1989) have found that selection and presentation order of exemplars can have a significant effect on a network's ability to learn and generalize.

the fifty-six issues had been held out of the estimation process and predicted by a discriminant model based on the other fifty-five observations. Hold-one-out sampling is unworkable for the back propagation method because of the lengthy training process required to produce a back propagation model. Use of hold-one-out sampling actually improves the generalization capability of any technique over the use of a single hold-out sample. With hold-one-out sampling, the classification technique has much more information on which to develop its model (fifty-five vs. sixteen observations in this case). For hold-one-out sampling results to be applicable to the population as a whole, the exemplars must be independent.

Surkhan & Singleton's results can be seen in Table 4. Clearly, the back propagation network's ability to form nonlinear combinations of the inputs allowed it to substantially outperform the linear discriminant analysis, even though the method of sampling favored the discriminant analysis. Surkhan & Singleton attribute the superior performance of the third back propagation network (7-10-5-2) to the descending number of neurons in the hidden layers. While this arrangement may be helpful given the very small training sample, it may also simply reflect the initial random conditions in the two networks before training. In any case, all three network architectures demonstrated considerable improvement over the standard technique usually applied to these problems.

**Table 4. Performance of Back Propagation and Discriminant Analysis
on Bond Rating Classification
(Results of Surkan & Singleton)**

	Hit-rate	
Classification Technique	Out of Sample	In Sample
Back Propagation (7-14-2 neurons)	65.0	75.0
Back Propagation (7-5-10-2 neurons)	80.0	100.0
Back Propagation (7-10-5-2 neurons)	88.0	100.0
Linear Discriminant Analysis	39.0	N.A.

Bankruptcy Prediction/Classification

A similar comparison between back propagation and discriminant analysis for bankruptcy prediction was performed by Odom & Sharda (1990). In this case, five financial ratios were used to analyze the bankruptcy/non-bankruptcy behavior of 129 firms over the 1975 to 1982 time

period. Odom & Sharda trained a single hidden layer back propagation network (five neurons in the hidden layer) to the observed behavior of seventy-five of the firms (thirty-eight bankruptcies, thirty-six non-bankruptcies). Fifty-five firms were held-out as a test sample (twenty-seven bankruptcies, twenty-eight non-bankruptcies). These same samples were analyzed using discriminant analysis. In addition to training on the sample with a near 50/50 distribution of non-bankruptcy/bankruptcy cases, the researchers randomly selected two smaller samples from the bankruptcy cases. This was done to more accurately reflect the actual distribution of bankruptcies for all firms in the United States. For one sample they selected nine of the thirty-eight bankruptcy cases from the full training sample (an 80/20 breakdown), and for the other they selected four (a 90/10 breakdown). The out-of-sample hit-rates for back propagation and discriminant analyses can be seen in Table 5.

**Table 5. Performance of Back Propagation and Discriminant Analysis
on Bankruptcy Prediction
(Results of Odom & Sharda)**

Classification Technique	Out-of-Sample Hit-rate		
	50/50 Training Sample	80/20 Training Sample	90/10 Training Sample
Back Propagation	81.8	78.2	81.8
Linear Discriminant Analysis	71.5	78.2	69.1

While these results do not show as large a difference between the two techniques, back propagation performed at least as well as discriminant analysis on all three cases. Despite the twenty-four hour training times required by back propagation on an IBM PC-XT, Odom & Sharda note that back propagation was robust to changing sample sizes. In addition, they observe that back propagation generally outperformed "a method that has become the rule in bankruptcy prediction".

Other Financial Applications

Kimoto, Asakawa, Yoda, and Takeoka (1990) have developed a modular system of back propagation networks and several training heuristics to predict good buy/sell decisions on the Tokyo stock exchange. They used the average of four back propagation networks, trained to different inputs such as the Dow Jones average, turnover, and foreign exchange rates, to predict a moving average of the changes in the Tokyo stock exchange average. This prediction was the

converted into a buy/sell signal. They also employed a hold-out sample to stop training when out-of-sample performance began to degrade; then they trained on all observations (including the hold-out-sample) the number of iterations suggested in the prior training run. The system was made adaptive by sliding a fixed size training window forward as more recent periods were to be predicted. While extensive test results were not published, the researchers found that the model performed better than a simple buy and hold strategy. They also found that the network had a higher correlation with the changes in the observed behavior of the exchange than a multiple regression analysis using the same inputs.

One of the most commercially successful neural network applications has been a mortgage underwriter (Collins, Ghosh, & Scofield, 1988) . In this application, a patented Reduced Coulomb Energy (RCE) network was used to emulate the mortgage underwriting decisions made by human underwriters. For a description of the RCE network see Reilly, Cooper, & Elbaum (1982) and Scofield (1988). Inputs to the network include the same financial information used by the human underwriters. In 1988, Collins et al. found the network to be slightly better than human underwriters. In particular the network was more consistent in its accept/deny decisions. The system has been in continual testing since that time in commercial environments. In 1990 the system was re-benchmarked (Reilly, Collins, Ghosh, & Scofield, 1990). Despite the fact that the system had not been retrained, it still performed well when compared to human experts. In a related area, the RCE network was applied to evaluate credit card applicants (Hiefietz, 1989).

Radar and Sonar Applications

While somewhat more removed from personnel analysis, classification of radar and sonar waveforms has been one of the most fruitful applications of neural networks. Although the problem domain is quite different, waveform classifiers must still map a high dimensional input vector into discreet classifications. Given the breadth of research and results in this field, a brief review is appropriate in this context.

Classification of Sonar Signals

One of the earliest successful applications of a back propagation network to a "real world" classification problem was performed by Gorman & Sejnowski (1988). They sought to distinguish between a metal cylinder and a rock on the floor of a pool using sonar waveforms. After training on 104 examples, the back propagation network was able to correctly classify 90.4% of another 104 examples from a hold-out sample. This compares with 82.7% for a nearest neighbor classifier and is almost identical to the performance of trained humans on the same task. Gorman & Sejnowski found that a network with twelve neurons in the hidden layer performed best, but that six were probably sufficient and increasing the number of neurons did

not significantly degrade the out-of-sample performance (the maximum they tested was twenty-four). Below six neurons, the network's performance degraded severely, indicating that the data contained some nonlinear relationships.

Classification of Radar Waveforms

A group of researchers has been working to identify the type of construction used to build bridge decks without physically sampling the decks (Vrckovnik, Chung, & Carter, 1990a;). The classes for the type of construction depend on such factors as whether the deck has a waterproof coating and how many layers of asphalt were laid down. In most tests, the researchers were trying to classify the waveforms among three different construction types. In all tests, samples of impulse radar waveforms from bridge decks of known construction were used. A very high input dimension of 140 time slices from the waveform were used. In early work, the researchers trained a back propagation network (140 inputs, 25 hidden, and 1 or two outputs) and a nearest neighbor classifier to a set of 1350 known waveforms. On this test, the network and nearest neighbor classifier performed comparably on a combination of all in- and out-of-sample exemplars[6]. Despite slow training for the back propagation network, the researchers found back propagation preferable in this case to nearest neighbor classification. Once training is complete, back propagation networks can perform classifications quickly. On the other hand, nearest neighbor classifiers require distance calculations between a new exemplar and all observations from the training sample.

In later research, Vrckovnik, Chung, & Carter (1990b) applied principal components analysis to the 140 dimensional input vector to extract the first fifteen principal components. The back propagation network was able to correctly classify 89.9% of all data (180 training samples and 720 hold-out samples) using the raw 140 dimension input. When the network was trained on the fifteen principal components, it obtained a 99.15% hit-rate on the same data set. Vrckovnik, Carter, & Kin (1990c) compared the back propagation network's performance to that of a radial basis function network (RBF)[7]. They found that the best RBF network could correctly classify 89.8% when using the 140 inputs and 99.7% when using the fifteen principal

---

[6]It is unknown why the researchers mix the training and hold-out samples when testing the network. This practice confounds the in- and out-of-sample measurements and makes it difficult to assess the generalization capability.

[7]For a description of RBFs (or localized receptive fields) see Moody & Darken (1988). These networks bear some resemblance to LVQ, or more properly the counterpropagtion network briefly mentioned in Section II. While several variations exist, they all use local receptors which preferentially respond to inputs which are close to the receptor's weights. The outputs of these receptors is then combined to form the networks output. These networks have been found to train as much as 1000 times faster than back propagation and usually produce comparable results. However, for a given prediction performance level, they generally require more training observations than back propagation.

15

component inputs. The performance on the combined training/testing data set was nearly identical for the back propagation and RBF networks (on both the 140 inputs and the fifteen principal components).

Other researchers have used back-propagation networks themselves to form nonlinear high-order features. The high-order features, similar to principal components, are generated by treating the inputs as training targets (Oja, 1982; Juell, Nygard, & Nagesh, 1988; Hrycej, 1990). Some researchers have also found that specific classification problems were easier to solve using these high-order features rather than the raw inputs. It is somewhat unusual that compressing the input dimensions would help back propagation in extracting the form of the input/class relationship. Since a feed-forward network can create any nonlinear relationship required, the removal of some information by compressing the input dimension would not seem to be helpful. Failure of the back propagation networks to perform well out-of-sample in these instances may be related to over-fitting because of the extra freedom afforded by the large number of inputs. Early research with personnel data suggest that over-training is a likely cause of the observed behavior. However, it is also possible local minima (or more likely long flat spots mistaken for local minima) may be trapping the learning process with the high dimensional inputs. Felton, Martin, Otto, & Hutchinson (1990) have found evidence of this latter behavior in high-dimensional character recognition tests.

A major problem with retaining the training inputs when testing the networks performance involves the local nature of nearest neighbor classifiers and RBF networks. If the training sample is tested in a nearest neighbor classifier, perfect classification will occur for all of the training cases. Each training exemplar will be closest to its own stored input vector and that vector will determine its predicted outcome. Barring multiple training exemplars with identical inputs, it is impossible for a nearest neighbor classifier to misclassify one of training exemplars. This condition holds to a lesser degree (or in some cases not all) for different variations of the RBF networks. For the network employed by Vrckovnik, the radial basis locations were drawn directly from exemplars in the training sample. This guarantees that when one of these exemplars is presented during testing, the basis location for the exemplar established during training will have its maximal response. This high response will heavily bias the network to classify the "testing" exemplar with training exemplar version of itself.

Orlando, Mann, & Haykin (1990) tested back propagation and LVQ networks against standard a Gaussian classifier (see Duda & Hart, 1973) in the classification of sea ice. The simple Gaussian classifier employed estimates a single multivariate Gaussian distribution around the sample mean of each class. A Bayesian decision rule is then applied to select the least expected loss class for any unknown exemplar. Orlando, et al. used a cross polarized radar producing only two intensities as the inputs to the classification techniques. They sought to classify the radar signals into four categories of sea ice: first year ice, multi-year ice, icebergs, and shadows cast by icebergs. Their results showed very little performance difference between the three classifiers. The out-of-sample hit-rates for the three techniques were: Gaussian 82.0%, back propagation 82.6%, and LVQ 81.7%. The reason for the failure of the networks to exceed the performance of the the simple Gaussian classifier could be seen when the

researchers graphed the decision regions for the three techniques. All three developed very similar decision regions and the regions were simple, contiguous, and similar to those expected of overlapping Gaussians. In this case, the data contained no nonlinearities or interactions which the network techniques could exploit. Still, they were able to form a "good" model of this problem; even if this model is directly estimable by a standard technique.

One of the first installed neural network applications involves the detection of plastic explosives in airline luggage (Shea & Lin 1989). A gamma ray generator and series of detectors to scan luggage is installed in a system much like the standard airport X-ray machines. Based on a set of features extracted from this system, the luggage is classified as to its likelihood of containing a bomb. During testing, it was found that a back propagation network performed superior to discriminant analysis on the task. Both methods can be "tuned" to determine how strong an output signal is required before a bomb is assumed. It was found the back propagation network, over the relevant operating range, always had a superior probability of detection for any given probability of a false alarm. When the system was installed by SAIC inc. and the FAA in JFK airport for on-site testing, this superior performance continued (Shea & Liu, 1990). The back propagation technique still displayed a superior detection to false alarm curve. In addition, when the "tuning" parameters were fixed at their desired level, the two classifiers performed almost identically at detecting explosives during operational testing (both had 98% detection rates). However, the back propagation network had a far superior false alarm rate: 7.8% compared to 11.6% for the discriminant classifier. This difference is very important for this problem; each of the false alarms represents a piece of luggage that must be hand searched. If FCC recommendations are implemented, these devices will be required equipment for international flights from all major U.S. airports.

Diagnostic Applications

Automotive Diagnostics

Marko, Feldkamp, & Pushkorius (1990) have reported on attempts to detect engine fault conditions based on information available from a car's electronic engine controller (EEC). Two data sets were available: direct readings from the 60-pin EEC and sequential data from the EEC's two-wire data control link (DCL). The first data set is very difficult to collect on-line but contains much information. The second data set is simple to collect but is much more difficult to analyze. It was found that expert engine diagnosticians could interpret the 60-pin EEC data and consistently recognize faults from this information. However, they could not express their techniques in a manner which allowed a rule based system to be constructed from their expertise. Using the DCL data alone, the experts were unable to assess fault conditions at all. The researchers tested several techniques on the classification problem: a binary tree

17

hyperplane classifier[8], a modification of the Gaussian classifier, a nearest neighbor classifier, a back propagation network, a single RCE network, and multiple RCE networks[9]. In all cases, Marko et al. tested the networks using the hold-one-out technique described earlier. For the 60-pin EEC data, the classifiers were trained to identify twenty-three fault conditions plus the standard no-fault condition. On the DCL data, the classifiers were trained to recognize seven major faults. The researchers reported error rates, but these have been transformed into the hit-rates displayed in Table 6.

All classifiers performed well on the "easier" ECC data, with the binary tree and nearest neighbor able to achieve perfect performance. The binary tree was considered superior because it requires considerably less storage and time to perform a classification. On the more difficult DCL data, the binary tree again exhibited the best performance. Back propagation was a close second and the other classifiers performed poorly. Because of their space and computational efficiency, Marko et al. spent much of their time developing and tuning the binary tree classifier. Conversely, their selection of the size and training parameters for the back propagation network was, in their own words, "rather arbitrary". Given their space and speed requirements, this was an optimal decision. It is unclear, however, how much impact this decision had on the relative performance of the techniques.

Electric Power System Security

Atlas, Cloe, Conner, El-Sharkawi, Marks, Muthusamy, & Barnard (1990) performed a series of three tests comparing back propagation networks and CART on "real-world" applications. The application considered here involved diagnosing (or predicting) when an a electrical power system was in a secure or unsecure state. The system is most efficient when in a near unsecure state, but is in brown-out or black-out when its state becomes unsecure. When selecting network (for back propagation) and tree (for CART) size, the researchers used a hold-out sample during initial training runs to determine a near-optimal size for generalization. Atlas et al. found that a back propagation network's out-of-sample prediction error rate on this task was 0.78%, compared to 1.46% for CART. This difference was found to be significant at the 99% level of confidence. The researchers also tried training (or
stimating) over various size training samples. In all cases, back propagation out-performed CART.

---

[8]A version of a hyperplane separator implemented as a neural network was employed by the researchers. For information on its implementation see Koutsougeras & Papachristou (1988). This classifier bears a strong resemblance to the CART technique discussed earlier.

[9]The multiple RCE network is a proprietary network of the Nestor Corp. employing a collection of single RCE networks.

**Table 6. Performance of Different Classifiers
for Automotive Engine Diagnostics
(Results of Marko, Feldkamp, & Puskorius)**

Classification Technique	Hold-One-Out Hit-rate	
	60-pin ECC Data	DCL Data
Binary Tree	100.0	92.0
Gaussian Adaptation	99.9	50.0
Nearest Neighbor	100.0	80.0
Back Propagation	-[1]	90.0
RCE (single)	97.5	56.0
RCE (multiple)	97.7	-[2]
[1]Back propagation was too compute intensive for hold-one-out sampling. [2]Due to scaling problems, multiple RCE networks were not applied		

Fault Detection in Chemical Plants

Hoskins, Kaliyur, and Himmelblau (1990) applied a back propagation network to a chemical plant diagnostic problem which had been "too difficult for traditional model-engineering and rule-based systems". They were analyzing the output of eighteen sensors, each taking eleven measurements, to isolate normal operating conditions and three faults: liquid flow resistance, gas flow resistance, and cooling motor resistance. They found that the network could learn to perfectly classify the faults when the training data set contained 10% noise. While they do not describe specific out-of-sample performance, they found the network could generalize its in-sample performance.

Defense Communication Satellite Diagnostics

One of the largest installed neural network systems is designed to automatically detect anomalies in defense communication satellite systems (DSCS) that cannot be detected with normal systems. The system (see Casselman & Acres, 1990) can diagnose thirteen problems

such as autotracking failure using nine feed-forward networks trained with back propagation. As of Casselman & Acres' 1990 paper, five of the systems are now installed at the DSCS operation center. They concluded that neural networks are now mature enough to apply to large scale diagnostic systems.

Phoneme Classification

While the problem domain of phoneme classification is quite different from personnel classification, the area has been extensively researched. Along with character recognition, speech recognition has probably received more neural network research attention than any other area. Phoneme classification problems involve the segmentation and mapping of voice waveforms onto a set of symbols representing specific phonemes. Often the waveforms are filtered through a variety of transformations such as fast fourier or gabor transforms. In general, the resulting classification task requires the development of highly nonlinear and often time dependent decision boundaries.

Alex Waibel (Waibel, 1988; Waibel, 1989, Hataoka & Waibel, 1990) has developed a "time aware" version of the back propagation network to perform phoneme classification. He has also explored methods of combining simple back propagation networks into larger systems using additional back propagation networks (or "connectionist glue"). To date these systems are still experimental; but Waibel has obtained results on test data comparable to state-of-the-art hidden Markhov models. This amounts to a 98.6% hit-rate when discriminating between six phonetically similar consonants (96% when all English consonant phonemes are considered).

In the area of vowel classification, Leung & Zue (1989) compared the results of several classifiers on a multi-speaker data set. They found that a simple back propagation network performed better out-of-sample than either K-nearest-neighbor or Gaussian classifiers. Atlas et al. (1990) extended their tests between CART and back propagation to speaker-independent vowel classification. Using sixty-four spectral coefficients from the waveform as inputs, they found that the back propagation network performed somewhat better than CART: 47.4% correct classification for back propagation vs. 46.4% for CART. This test was based on a very limited window of information from the speech signal. When trained individuals were provided with the same window of information, they could only manage a 51% classification rate.

Other Classification Applications and Comparisons

Many other classification problems have been attempted with neural networks ranging from handwritten character recognition to classification of insect courtship songs (Neumann, Wheeler, Burnside, Bernstein, & Hall, 1990). A brief review of some of these projects will demonstrate the breadth of classification applications and relative performance of neural networks.

20

Medical Diagnosis

Several researchers have examined the use of neural networks in medical diagnosis. Some of the earliest results were reported by Donald Specht (1967) using a non-neural network implementation of the algorithms used in his PNN. In this study, Specht analyzed 312 vector cardiograms along a forty-six dimensional input vector. Training with his algorithm on 249 exemplars, he was able to correctly generate a "normal heart" diagnosis for all the normal hearts in the sixty-three case hold-out sample. A nearest neighbor classifier was correct on 97% of these cases. The precursor to the PNN correctly classified 90% of the abnormal hearts in the hold-out sample; while the nearest neighbor classifier managed 74%.

Star Pattern Recognition

Researchers at the Jet Propulsion Laboratory (JPL) have been experimenting with neural networks to recognize star field patterns from unmanned inter-planetary satellites. Self alignment of the communication antennae of these satellites is critical when the satellite becomes mis-oriented during operation. Currently, the realignment process requires a very time consuming search process because the satellite does not know its orientation. The best current technique for determining this orientation involves comparing the stars in a specific field with those in a catalogue using standard search techniques. While the test system performs adequately, 99% success rate with one second search, it requires 650K of memory and a microprocessor. This system is much too large to fit within the constraints of a satellite. A back propagation network has been trained to perform the same task in about one tenth of a second and using only 12K of memory.

Character Recognition

The results of Denker, Gardner, Graf, Henderson, Howard, Hubbard, Jackel, Baird, & Guyon (1989) are typical of the many projects in the character recognition fields. This group sought to recognize handwritten numerals from zip codes. The sample consisted of 10,000 zip codes which had been digitized by the Postal Service from envelopes. Standard techniques and heuristics were used to scale, skeletonize, and detect features in the scanned images. This information was then used as input to three classifiers: K-nearest neighbor, Parzen windows [10], and a version of back propagation (forty neurons in one hidden layer). The classes consisted of the ten digits -- zero to nine. They found that the back propagation network performed better than the other two techniques. If the network were allowed to reject 14% of a test sample as unclassifiable, it could obtain 99% correct classification on the remainder of the sample. When forced to classify all exemplars from the test sample, the network correctly classified 84% of the sample. As an indication of the importance of pre-processing, when classification was attempted using the raw 256 bit input vectors, all of the classifiers performance fell by about

---

[10]Parzen windows are used to estimate probability densities and are somewhat similar to the PNN (see Duda & Hart, 1973 for details).

80%. While such factors as translation and rotation invariance are not usually important in personnel data, these results indicate that appropriate pre-processing can have a substantial impact on the performance of back propagation and other classifiers.

PREDICTION APPLICATIONS

A second broad area of interest to personnel researchers and planners is prediction or forecasting. The distinction between prediction and many of the other neural network application areas is somewhat arbitrary. While the classification of an airman among the possible reenlist/separate/extend decisions may be considered prediction, in this section prediction will be restricted to the projection of real-valued outputs. In the personnel area, these types of prediction models are more applicable to overall inventory flow projections than the classification methods just discussed. As with the classification techniques, no research has yet been published where neural networks are applied to prediction in the personnel area.

Chaotic Time Series

The prediction of chaotic time series has served much the same purpose as classification tests on known or contrived problems. The most common time series used in these prediction tests is the Mackey-Glass equation. This equation is self-iterated, one-dimensional, and belongs to the class of functions which produce deterministic chaos. While the outputs are completely deterministic and based entirely on the equation, they have the appearance of randomness and are extremely sensitive to initial conditions. Points which start out very close together, have time series paths which diverge considerable. Despite its one-dimensional nature, the Mackey-Glass equation has been extremely resistent to projection with traditional techniques. As Moody & Darken (1988) note, auto-regressive and polynomial expansion techniques generally fail on this data set. That is, the normalized prediction error[11] of the techniques is usually close to 1.0 (or no better than predicting the mean of the estimation data set).

Moody and Darken tested two network architectures on the Mackey-Glass equation: RBF[12] and back propagation. They used an auto-regressive set of inputs containing the last four values of the one dimensional equation (X_{t-1} to X_{t-4}) to predict the current value of the function (X_t). They found the performance of the two networks was very similar. The out-of-

---

[11]The normalized prediction error is just a scaling of the root mean square error (RMSE): RMSE / (standard deviation of the series). This scaling generally keeps the result between zero and one where zero is perfect prediction and one is no better than the mean of the series.

[12]The authors used the term localized receptive field, but radial basis function (RBF) will be used here to maintain consistency with earlier discussions. In fact, they used several adaptive techniques similar to those used in LVQ to generate the receptive fields.

sample normalized prediction error for back propagation was 0.06 and for RBF was 0.08. Given the complexity of the problem, these are extremely good projection fits. Moody and Darken found that the RBF network could be trained about 1000 times faster than a back propagation network. However, they also found that more training exemplars were required by the RBF network to obtain the same prediction accuracy as the back propagation network. Since the RBF network computes approximations to the underlying model using local basis functions, this result is not surprising.

Several other researchers have approached the Mackey-Glass equation with neural networks. Farmer & Sidorowich (1987) pioneered the use of back propagation on this equation. Their results were similar to those of Moody & Darken. Jones, Lee, Barnes, Flake, Lee, Lewis, & Qian (1990) recently extended the radial basis function network to reduce training requirements. They were able to obtain very good results using only five basis functions and ten training exemplars (less than 10% error for the worst six-step forward prediction). Walter, Ritter, & Schulten (1990) successfully applied Kohonen's self-organizing map to the problem. (This architecture is related to the LVQ learning algorithm and forms topologically correct low dimensional maps from high dimensional feature sets; see Kohonen, 1984). They were able to generate thirty-eight step predictions which had a .97 correlation with the actual results from the Mackey-Glass equation.

Electric Power Load Forecasting

Atlas et al. (1990) continued their comparison of CART and back propagation with the "real-world" example of electric power load forecasting for the Seattle/Tacoma region. Training was done on fifty-three days of hourly load and temperature. The resulting models were then used to forecast four days of out-of-sample load. They found that the average error rate for back propagation was 1.39% and the rate for CART was 2.86%. While this difference is not statistically different, the back propagation performed better over the test period than the model currently employed by the utility.

Expert System Solar Flare Forecasting

Bradshaw, Fozzard, & Ceci (1989) compared the ability of a back propagation network to project the probability of experiencing three types of solar flares over a twenty-four hour period. The network was trained was trained on 500 observations and tested on an independent sample of 500 observations. The network was found to have slightly better performance than an expert system using the same input information. Bradshaw et al. note that the expert system took over one man-year to develop and contained over 700 rules while the network was developed in one week with a simple simulator. In addition, the expert system requires about five minutes of processing to produce a prediction while the neural network takes less than one second. However, the expert system can explain its predictions and the neural network cannot.

Stock Market Prediction

A test of the efficient markets hypothesis from economics was performed by Halbert White (1988) using daily stock returns for IBM. As part of this test, White developed a linear auto-regressive model and an auto-regressive model using a back propagation neural network. When the linear auto-regressive model was estimated on 1,000 daily returns, the R-squared for the equation was 0.008. This estimate is not significant at the 10% level and thus it is doubtful that this model has captured any true structure in the series of prices. On the other hand, the back propagation model had an in-sample R-squared of 0.175 which on the surface is very impressive. However, when the back propagation model was tested out-of-sample (both prior to and after the estimation sample), the correlation of the predictions with the actual returns was very small and insignificant. In fact, in one case, the predictions and actual returns were negatively correlated. In this case, the back propagation network failed to generalize out-of-sample. As White notes, the network may have discovered fleeting structures. These would have been actual features of the underlying process during the sample time period, but they were not part of the underlying process over the test sample time periods. More likely, the back propagation network simply over-fitted the training observations which degraded its out-of-sample performance. On the other hand, the prices may truly obey the efficient markets hypothesis and are thus unpredictable from observations on past behavior. One would expect this to be a particularly noisy problem domain and White did not attempt to use any techniques to improve the generalization ability of back propagation.

CONTROL APPLICATIONS

Control of systems, particularly in an adaptive environment, are another domain that has received the attention of network researchers. Many aspects of the Air Force personnel system can be viewed from the perspective of adaptive control. Managing the flow of personnel to and from bases requires controls based on preserving or improving the readiness of the force. Likewise, promotion and accession flows can be considered from this perspective. Current neural network applications in control tend to be quite different from the control of personnel flows. Even so, a brief overview of these applications demonstrates some techniques that may be applicable in the manpower area.

Aircraft Control

Josin (1990) has reported on a neural network autopilot being developed for NASA. Currently, the controller is operating with a software simulator of a high performance aircraft. Using a small back propagation network, the controller can be directed to attain and maintain

level flight at a prescribed height above a target. Given the current height, horizontal velocity, horizontal acceleration, and aircraft pitch, the network controls changes in "stick" position. The network controller is able to maintain a much tighter band around the target height than conventional autopilots.

Robotics Control

Many researchers are evaluating the use of neural networks for control of robotic arms. These applications typically focus on computing the inverse transformation that allows a request in X-Y or X-Y-Z coordinate space to be converted into angles between the components of a robot arm. This inverse transformation can be solved directly for arms with few degrees of freedom (joints), but becomes intractable when the arm contains four or five degrees of freedom. Several investigators (Artego & Bravo, 1990; Josin, 1988) have employed back propagation networks to "learn" the X-Y coordinate to arm-angle mapping for two degree of freedom arms. Josin, in particular, has shown that the network can successfully generalize the transformation with as few as three training exemplars. Control of a five degree of freedom industrial robot has also been demonstrated by Josin (1989). In this case, the standard controller failed to provide enough precision in positioning the arm. A back propagation network was trained as an adjunct to the standard controller which then served to adjust for the errors made by the original controller.

Automobile Control

Another group (Shepanski & Macy, 1987) trained a neural network to "drive" a simulation of a car. The setting for the simulation was a two-lane road with turns and other cars travelling at random speeds. A back propagation network was provided with information on the distance and relative speed of the other cars, current car orientation, current lane, and lane curvature. The control movements of a human subject supplied with the same information were supplied as target outputs to the network. The network learned to perform the necessary speed and steering angle adjustments and could safely navigate the course after training. In addition, the network assumed the driving characteristics of the human subject who supplied training input. If the subject was reckless, the trained network exhibited the same "reckless" characteristic. This experiment demonstrated the "master-slave" learning technique; and also demonstrated the ability of neural networks to mimic some human control behaviors.

Other Control

Other researchers have examined the performance of neural networks for various control tasks. Blumenfeld (1990) trained a back propagation network with simple recurrent connections (Elman, 1988) to control insulin dosages and maintain patient glucose levels. Porcino and

Collins successfully applied the adaptive critic network of Barto & Sutton (1983) to the guidance of free-swimming submersibles. As with classification, many other examples of applying neural networks to control problems have been documented.

COGNITIVE APPLICATIONS

A very different area of neural network research involves cognitive functions such as planning, language comprehension, and expert behaviors. While these problems are quite different from those considered thus far, they are related to several application domains in the personnel management area. In particular, the ability reproduce decisions and behaviors of subject matter experts could be of use in many areas. A network could analyze a set of information and recommend several courses of action which match well with past actions taken by domain experts. While not directly related to the current research, some neural network paradigms may be able to assist in the mapping of job and task descriptions between disparate databases.

Grammar and Word Comprehension

Elman (1989 & 1990) has examined the ability of a simple, recurrent back propagation network to learn and reproduce valid examples from a fixed grammar. By simply training the network to predict the next word in a sentence, Ellman's network was able to learn valid use of the parts of speech. The network could also recognize and reproduce proper use of plurals. In addition, fine distinctions could be made about specific word choice based on the prior context of the sentence. Using the same architecture, another group of researchers (Servan-Schreiber, Cleeremans, & McClelland, 1989) trained a network to recognize valid strings produced by a finite state automaton. Both of these results, demonstrate the ability of simple recurrent networks to extract useful contextual information from a sequence of events.

Using a very different architecture, Ritter and Kohonen (1990) were able to generate self-organizing feature maps of common words from a corpus of simple sentences. These two-dimensional maps were able to represent the relative similarity or difference between words based solely on their context in the corpus of sentences. The maps grouped parts of speech such as noun or verb together in the map. Finer word distinctions also developed localized representations. Active verbs such as hit and run clustered together, as did antonyms such as far and near. Like Elman's simple recurrent networks, the feature maps proved capable of extracting meaningful concepts from the context of words in valid sentences.

Selection of Aircraft Combat Maneuvers

McMahon (1990) has compared the performance of a back propagation network to an expert system in selecting maneuvers during simulated aircraft combat. The neural network was derived from, and compared against, an existing air combat expert system -- the Air Combat Expert Simulation (ACES). ACES consists of thirty-eight production rules which produce twelve offensive and five defensive maneuvers. The system is supplied with twelve inputs representing the current battle situation (distance to enemy plane, relative position, etc.) on which to base its decision. McMahon trained a back propagation network using the thirty-eight production rules as prototypes to supply network inputs and target. Each of the production rules selected a specific maneuver if a sub-set of the twelve inputs matched particular ranges in the rule. Where an input was not used by a production rule, McMahon set the input to a random value when training the network. For validation purposes, forty combat scenarios were presented to a group of expert fighter pilots; and, the maneuvers chosen by the pilots were used to assess the performance of the two systems. The trained neural network, Tactical Air Combat Intelligent Trainer (TACIT), outperformed the ACES system despite having only the ACES production rules as training input. ACES agreed with the pilots' maneuver selection on 25% of the scenarios (ten of forty); TACIT agreed with the pilots on 67% of the scenarios (27 of 40). Because both systems were based on the same production rules, the network's superior performance is based on a better resolution of mutually consistent production rules. In some cases, more than one production rule is valid for a given set of inputs. The resolution strategy in ACES to select among these competing rules proved inferior to resolutions made by the neural network.

SUMMARY AND CONCLUSIONS FROM LITERATURE REVIEW

Various neural network architectures have been tested against standard statistical and expert system techniques in many problem domains. Some of the tests involved analyses of contrived problems and data sets while others were based on a wide variety of real-world problems. In the most of these tests, neural networks have been found to perform better than the traditional techniques; and, in virtually all cases, the network solutions at least equalled the traditional solutions. Neural network techniques have been successfully applied to classification problems ranging from financial bond rating and heart ailment diagnosis to radar waveform analysis and automotive engine fault detection. Many other applications have been tested in control, prediction, and cognitive tasks. Despite the success of networks on test problems, few commercial applications have yet been fielded. In addition, most of the current applications involve relatively small and well defined problem domains. Classification remains the most mature area of analysis, followed by control and prediction.

Theoretical results have demonstrated some of the reasons for the comparative success of neural networks. In particular, feed-forward networks have been shown to be capable of

approximating any continuous functional mapping. They are also capable of performing a nonlinear analogue of discriminant analysis. These behaviors are beyond most statistical techniques and thus provide the networks with added capabilities. These theoretical advantages were manifest in the empirical results described above. Neural network techniques typically performed better than traditional techniques on both contrived and real-world problems. In general, the networks demonstrated superior in- and out-of-sample performance when compared to discriminant analysis, regression, nearest neighbor analysis, K-means, CART, and some expert systems. Preliminary work by the authors using back propagation, PNN, and LVQ networks for personnel analysis has shown some improvement over logit and probit models.

While theoretical analysis has demonstrated some important capabilities of properly trained neural networks, less is known about the complex dynamic training process. Convergence to a global optimum is not generally guaranteed and some problems have been proven to contain local minima. However, in empirical tests against traditional techniques, this potential problem has not significantly affected the relative performance of the networks.

Some researchers found the networks could be applied directly and still perform well out-of-sample. Others found that the size of the networks required tuning on each specific problem to obtain good generalization properties. This tuning process is similar to specification searches for standard modeling techniques (although the tuning is much simpler). Currently, no theoretical or definitive empirical results provide specific guidance on training practices to maximize the ability of a network to generalize (perform well out-of-sample). The applications reported above were chosen because they provided some of the most complete in- and out-of-sample performance comparisons. Many other studies ignored out-of-sample performance which severely biases any empirical comparisons in favor of the highly flexible neural networks. Preliminary tests of neural network personnel models should provide even more complete in- and out-of-sample testing (see Appendix A). Identification of training processes and network architectures that improve generalization is expected to be an important aspect of applying neural networks to personnel modeling.

Overall, the success of neural networks in application domains similar to many personnel modeling problems is encouraging. Empirical results have often been impressive when compared to standard modeling and classification techniques. While some important theoretical results have been proven, training dynamics and factors contributing to generalization are not well understood. Most of the current applications are in small, well defined problem domains. The computational requirements of simulating neural networks on serial computers also places limits on the size of problems which can be readily addressed. While hardware accelerators and the recent development of neural network semiconductor chips will allow much larger problems to be addressed, these more expensive options should be delayed until the networks have proven to be capable on smaller personnel problems. Preliminary personnel models should be tested on reasonably small problems to allow thorough analysis of the neural network results and comparisons with standard techniques and existing models.

28

REFERENCES

Ackley, D.H., Hinton, G.E., & Sejnousti, TJ (1985). A learning algorithm for Boltzman machines. *Cognititive Science, 2* - 147-169.

Arteago-Bravo, (1990). Multilayer back propagation network for learning the forward and inverse kinematics equations. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS, 2,* Caudill, M (ed.), II - 319.

Atlas, L., Cole, R., Conner, T., El-Sharkaui, M., Marks, R.J. II, Muthusamy, Y, Barnard, E. (1990). Performance comparisons between backpropagation networks and classification trees on three real-world applications. *Neural Information Processing Systems 2,* Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 623-629.

Barto, G., Sutton, S., Anderson, C. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics,* SMC-13, 834-846.

Blumenfeld, (1990). A connectionist approach to the processing of time dependent medical parameters. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS, 1,* Caudill, M (ed.), II - 575.

Bradshaw, G., Fozzard, R., Ceci, L. (1988). A connectionist expert system that actually works. *Neural Information Processing Systems 1,* Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 248-255.

Casselman & Acres, (1990). DASA/LARS, a large diagnostic system using neural networks. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS, 1,* Caudill, M (ed.), II - 539.

Collins, E., Ghosh, S, Scofield, C. (1988). An application of multiple neural network learning system to emulation of mortgage underwriting judgement. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988, 2,* II - 459-466.

de Bollivier, M., Galliari, P., Thiria, S. (1990). Cooperation of neural networks for robust classification. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS, 1,* I - 113-120.

Denker, J.S., Gardner, W.R., Graf, H.P., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., Baird, H.S., & Guyon, I. (1989). Neural network recognizer for hand-written zip code digits. *Neural Information Processing Systems 2,* Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 323-331.

Devijver, P. & Kittler, J. (1982). *Statistical Patter Recognition*. London: Prentice Hall.

Duda, R.O., & Hart, P.E. (1973). *Pattern classification and scene analysis*. New York: John Wiley & Sons.

Elman, J.L., (1989). *Representation and structure in connection models* (CRL-8903). University of California at S. n Diego: Center for Research in Language.

Elman, J.L., (1990). Finding structure in time. *Cognitive Science*. 14, 179-211.

Farmer, J.D. & Sidokowich, F., (1987). Predicting chaotic time series. *Physical Review Letters*, 59, 845-854.

Felten, E., Martin, O., Otto, S., & Hutchinson, J. (1990). Multi-scale training of a large backpropagation net. *Biological Cybernetics*, 62(6), 503-509.

Gallinari, P, Thiria, S., Soulie, F.F. (1988). Multilayer perceptions and data analysis. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988*, 1, I - 391-399.

Gorman, R.P., & Sejnowski, T.J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. Neural Networks, 1(1), 75-89.

Hancock, P.J.B (1988). Data representation in neural nets: an emprical study. *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University*, Touretski, D., Hinton, G., & Sejnowski, T. (eds.), San Mateo, CA: Morgan Kaufmann Publishers, 11-20.

Hataoka, N., & Waibel, A.H., (1990). Speaker independent phoneme recognition on TIMIT database using integrated time-delay neural networks (TDNNs). *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 57-62.

Hoskins, J.C. (1989). *Speeding up artificial neural networks in the "real" world* (pre-print). Austin, TX: Microelectronics and Computer Technology Corporation.

Hoskins, J.C., Kaliyur, K.M., Himmelblau, D.M., (1990). Insipient fault detection & diagnosis using artificial neural networks. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 81-86.

Hrycej, T. (1990). Self-organization by delta rule. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 2, II - 307-312.

Huang, W.Y. & Lippman, R.P. (1987). Comparisons between neural network and conventional classifiers. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987,* 4, IV - 485-493.

Hush, D.R., & Salas, J.M., (1990). A performance of neural network classifyers for the 1-class classifyer problem, *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS,* 1, Caudill, M (ed.), I - 396-399.

Jones, R.D., Lee, Y.C., Barnes, C.W., Flake, G.W., Lee, K., Lewis, P.S., & Qian, S. (1990). Function approximation and time series prediction with neural networks. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS,* 1, I - 649-665.

Josin, G., (1990). Development of a neural network autopilot model ford high performance aircraft. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS,* 1, Caudill, M (ed.), II - 547.

Josin, G., (1989). An application of neural networks to the guidance of free-swimming submersibles. Invited tutorial, ORS/TIMS/ORSA Joint National Meeting, Vancouver, Canada, May 8-10, 1989.

Josin, G., (1988). Neural-space generalization of a topological transformation. *Biological Cybernetics,* 59, 283-290.

Juell, P., Nygard, K., & Nagesh, K. (1989). Multiple neural networks for selecting a problem solving technique. Unpublished report, Department of Computer Science and Operations Research, North Dakota State University.

Kimoto, T., Asakawa, K., Yoda, M., & Takewka, M. (1990). Stock market prediction with modular neural networks. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS,* 1, I - 1-6.

Kohonen, T., Barna, G., Chrisley, R (1988). Statistical pattern recognition with neural networks. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988,* 1, I - 61-88.

Koutsougeras, C., & Papachristov, (1988). Training of a neural network for data class based on an entopy measure. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988,* 1, I - 247-252.

Lang, K.J., & Witbrock, M.J., (1988). Learning to tell two spirals apart. *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University,* Touretski, D., Hinton, G., & Sejnowski, T. (eds.), San Mateo, CA: Morgan Kaufmann Publishers, 52-59.

Leung, H.C., & Zue, V.W., (1988). Applications of error back propagation to phonetic classification. *Neural Information Processing Systems 1*, Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 206-214.

Lippman, R.P. (1987). An introduction to computing with neural networks. IEEE ASSP Magazine, 4, 4-22.

Maddala, G. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.

Marko, K., Feldkamp, L., & Puskorius, G. (1990). *Automotive diagnostics using trainable classifiers: statistical testing and paradigm selection. International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 33-38.

Moody, J., & Darken, C. (1988). Learning with localized receptive fields. *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University*, Touretski, D., Hinton, G., & Sejnowski, T. (eds.), San Mateo, CA: Morgan Kaufmann Publishers, 133.

McMahon, D.C., (1990). A neural network trained to select aircraft maneuvers during air combat: a comparison of network & rule based performance. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 107-112.

Neumann, E.K., Wheeler, D.A., Burnside, J.W., Bernstein, A.S., Hall, C. (1990). A technique for the classification and analysis of insect courtship song. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS*, 1, Caudill, M (ed.), II - 257.

Odom, M.D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 2, II - 163-168.

Oja, E. (1982). A simplified neuron model as a principle component analyzer, *Journal of Mathematical Biology*, 15, 267-273.

Orlando, T., Mann, R., & Haykin, S. (1989). *International Joint Conference on Neural Networks, Washington D.C., June 18-22, 1989, IEEE & INNS*, 2, II - 263-266.

Ritter, H., Kohenen, T., (1990). Learning "semantotopic maps" from context. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS*, 1, Caudill, M (ed.), I - 23-26.

Servan-Schreiber, P, Cleermans, A., & McClelland, J. (1988). Learning sequential structure in simple recurrent networks. *Neural Information Processing Systems 1*, Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 643.

Shea, P.M., & Liu, F. (1990). Operational experience with a neural network in the detection of explosives in checked airline luggage. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 2, II - 175-178.

Shea, P., & Lin, V. (1989). Detection of explosives in checked airline baggage using an artificial neural system. *International Joint Conference on Neural Networks, Washington D.C., June 18-22, 1989, IEEE & INNS*, 2, II - 31-34.

Shepanski, J.F., & Macy, S.A., (1987). Manual training of autonomovs systems based on artificial neural networks. *IEEE First International Conference on Neural Networks, San Diego, California, June 21-24, 1987*, 4, IV - 697-704.

Specht, D.F. (1967). Vectorcardiographic diagnosis using the polynomial discriminant method of pattern recognition. *IEEE Transactions on Bio-medical Engineering*, 14, 96-95.

Surkan, A.J., & Singleton, J.C. (1990). Neural networks for bond rating improved by multiple hidden layers. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 2, II - 157-162.

Vrckovnik, G., Chung, T., & Carter, C.R. (1990). An application of neural network to impulse radar waveforms from asphalt-covered bridge decks. *International Joint Conference on Neural Networks, January 15-19, 1990, Washington D.C., IEEE & INNS*, 2, Caudill, M (ed.), II - 453.

Vrckovnik, G., Carter, C., Hay Kin, S. (1990). Radial basis function classification of impulse radar waveforms. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 1.

Waibel, A. (1989). Consonant recognition by modular construction of large phonetic time-delay neural networks. *Neural Information Processing Systems 1*, Touretzky, D.S. (ed.), San Mateo, CA: Morgan Kaufmann Publishers, 215-223.

Waibel, A., (1988). Connectionist glue: modular design of neural speech systems. *Proceedings of the 1988 Connectionist Models Summer School, June 17-26, 1988, Carnegie Mellon University*, Touretski, D., Hinton, G., & Sejnowski, T. (eds.), San Mateo, CA: Morgan Kaufmann Publishers, 417-425.

Walter, J., Ritter, H., Schulten, K. (1990). Non-linear prediction with self-organizing maps. *International Joint Conference on Neural Networks, San Diego, California, June 17-21, 1990, IEEE & INNS*, 1, I - 589-584.

White, H. (1988). Economic prediction using neural networks: the case of IBM daily stock returns. *IEEE International Conference on Neural Networks, San Diego, California, July, 1988,* 2, II - 451-458.